Cognitive ergonomics for data analysis. Experimental study of cognitive limitations in a data-based judgement task

Kalakoski, Virpi

2019-10-03

Kalakoski , V , Henelius , A , Oikarinen , E , Ukkonen , A & Puolamäki , K 2019 , ' Cognitive ergonomics for data analysis. Experimental study of cognitive limitations in a data-based judgement task ' , Behaviour & Information Technology , vol. 38 , no. 10 , pp. 1038-1047 . https://doi.org/10.1080/014

http://hdl.handle.net/10138/306064
https://doi.org/10.1080/0144929X.2019.1657181

Behaviour & Information Technology

ISSN: 0144-929X (Print) 1362-3001 (Online) Journal homepage: https://www.tandfonline.com/loi/tbit20

Cognitive ergonomics for data analysis. Experimental study of cognitive limitations in a data-based judgement task

Virpi Kalakoski, Andreas Henelius, Emilia Oikarinen, Antti Ukkonen & Kai Puolamäki

To cite this article: Virpi Kalakoski, Andreas Henelius, Emilia Oikarinen, Antti Ukkonen & Kai Puolamäki (2019): Cognitive ergonomics for data analysis. Experimental study of cognitive limitations in a data-based judgement task, Behaviour & Information Technology, DOI: 10.1080/0144929X.2019.1657181

To link to this article: https://doi.org/10.1080/0144929X.2019.1657181

Published online: 24 Aug 2019.

BEHAVIOUR & INFORMATION TECHNOLOGY
https://doi.org/10.1080/0144929X.2019.1657181

Cognitive ergonomics for data analysis. Experimental study of cognitive limitations in a data-based judgement task

Virpi Kalakoski[a], Andreas Henelius[b], Emilia Oikarinen[b], Antti Ukkonen[b] and Kai Puolamäki[b]

[a]Finnish Institute of Occupational Health, Helsinki, Finland; [b]Department of Computer Science, University of Helsinki, Helsinki, Finland

ABSTRACT
Today's ever-increasing amount of data places new demands on cognitive ergonomics and requires new design ideas to ensure successful human–data interaction. Our aim was to identify the cognitive factors that must be considered when designing systems to improve decision-making based on large amounts of data. We constructed a task that simulates the typical cognitive demands people encounter in data analysis situations. We demonstrate some essential cognitive limitations using a behavioural experiment with 20 participants. The studied task presented the participants with critical and noncritical attributes that contained information on two groups of people. They had to select the response option (group) with the higher level of critical attributes. The results showed that accuracy of judgement decreased as the amount of information increased, and that judgement was affected by irrelevant information. Our results thus demonstrate critical cognitive limitations when people utilise data and suggest a cognitive bias in data-based decision-making. Therefore, when designing for cognition, we should consider the human cognitive limitations that are manifested in a data analysis context. Furthermore, we need general cognitive ergonomic guidelines for design that support the utilisation of data and improve data-based decision-making.

ARTICLE HISTORY
Received 11 July 2019
Accepted 13 August 2019

KEYWORDS
Human–data interaction; cognition; decision-making; cognitive ergonomics

1. Introduction

In this era of digitalisation, humans need to increasingly make decisions that rely on data that is analysed and summarised by algorithms and systems. Data-based decision-making may appear simple on the surface: systems such as search engines provide the most important parts of the data for experts and laymen interacting with the system to enable them to make rational decisions based on the evidence presented. However, previous research on human engineering has already demonstrated that several factors related to the usability of the system, sociotechnical context, and limitations of human cognition affect whether human–computer interaction is effortless and error-free (Wickens, Gordon, and Liu 2004). It is evident that the increasing amount of data available for open usage raises similar human factor issues, but it also raises new questions that need to be answered to ensure successful human–data interaction (Hibbard and Peters 2003).

Our focus is on the cognitive factors in human–data interaction. Ergonomic (or human factors) practices aim to ensure 'appropriate interaction between work, product and environment, and human needs, capabilities and limitations' (HFES 2019). Cognitive ergonomics is a domain of specialisation within ergonomics that concerns 'mental processes, such as perception, memory, reasoning, and motor response, as they affect interactions among humans and other elements of a system' (IEA 2019). When studying the factors that predict successful human–data interaction, it is important to understand how the cognitive aspects of both the human and system parts of the interaction constrain data-based decision-making.

Although the amount of data has increased and the methods for analysing them have advanced considerably in recent years, the basic cognitive ability of human beings has not developed in the same way. On the one hand, human decision-making is still limited by, for example, our capacity to rehearse and process information in short-term working memory (Baddeley and Hitch 1974; Cowan 2001), our inclination to better recall the first and last items from serially presented information (Glanzer and Cunitz 1966), and the various cognitive tendencies that bias our decisions (Tversky and Kahneman 1981). On the other hand, although the human ability to learn is exceptional, developing expert-level knowledge and skills is time-consuming (Ericsson and Lehmann 1996), and we seldom acquire

CONTACT Virpi Kalakoski virpi.kalakoski@ttl.fi Finnish Institute of Occupational Health, P.O. Box 40, FI-00032 Työterveyslaitos

exceptional or even adequate skill levels in a wide range of domains. For example, domain experts are rarely experts in data analysis, and they often have little knowhow or control of the systems used to analyse and summarise their data. In sum, when studying human–data interaction, it is important to identify relevant cognitive functions and to apply the findings of the related experimental and applied cognitive psychology studies (Kalakoski 2016).

Moreover, to understand human–data interaction, we also need to consider the system part of the interaction, that is, how data systems constrain the interaction. Although advances in computation and data analysis make it possible to differentiate complex patterns from large heterogeneous data sets, a query may appear fundamentally different when its parameters or filtering criteria are changed. As a result, we live in 'filter bubbles' created by information-processing systems (Pariser 2011), and large parts of important, significant or novel views of the data may remain unnoticed. Therefore, despite the availability of huge amounts of data, we are not able to thoroughly utilise the information or the evidence present in them. In sum, not only human cognitive limitations, but also the factors inherent in analysing large data sets can affect the success of human–data interaction. Our decisions may be founded on narrow views provided by the system, which exclude the critical factors necessary for making good, informed decisions (Hibbard and Peters 2003).

In naturalistic data-based decision-making situations, datasets are often complex and the problems ill-defined, that is, it is not clear if the chosen interpretation and decision is correct. In our experimental study, we demonstrate some cognitive phenomena that are relevant to making judgements in data-based decision-making tasks. We developed a task that represents central cognitive demands in human–data interaction. By manipulating the cognitive demands of the task, we investigated which factors affect decision-making task performance. Our practical objective was to identify the factors that need attention when aiming for cognitively sound designs. The results suggest some general guidelines for how to take cognition into account when designing data analysis systems.

2. Decision-making in applied contexts

Decision-making has been experimentally studied using various tasks and paradigms, and from many theoretical approaches. Since data-based judgement requires numerous cognitive processes, no single paradigm or theoretical approach can identify it is as a widely recognised phenomenon. However, several frameworks and cognitive phenomena are relevant. This notion is in line with the claim by Logie, Trawley, and Law (2011) that, in the context of multitasking, current cognitive theories can address specific components of the cognitive system rather than how the system performs complex tasks. Therefore, understanding everyday cognitive tasks requires new paradigms and theories that can handle various cognitive functions operating in concert rather than in isolation (Logie, Trawley, and Law 2011). We proceed with this line of thought in our study and, rather than studying a specific cognitive function or theoretical assumption, we aim to understand the cognitive demands and processes that are crucial in the applied context of data-based decision-making.

Research on decision-making in applied contexts offers an increasing number and scope of studies. The paradigms and frameworks, however, have been developed for specific contexts and are relevant to specific fields. For example, in the naturalistic decision-making research approach (Klein 2008), the models and methods apply to dynamic, continually changing conditions under which knowledgeable people work on ill-defined goals and ill-structured tasks (Klein et al. 1993). Similarly, the useful concept of situational awareness (Endsley 1995) combines several cognitive functions that are relevant in demanding dynamic tasks such as air traffic control (Endsley and Rodgers 1994). However, the concept and framework of this approach refer to a specific combination of cognitive functions and task demands and are not directly applicable to other contexts. Data-based decision-making is a complex everyday task but does not carry the features relevant to the above-mentioned frameworks of complex decision-making.

However, decision-making has also been studied in specific contexts that are applicable in human–data interaction. For example, the cognitive demands of using big data are similar to cases in which the consumer's health care decisions are supported by the use of information (Hibbard and Peters 2003) or when an auditor is provided with accounting information on the economic reality of the company (Lau 2008). In these examples, making informed decisions is a cognitively complex task that involves several cognitive steps, such as taking in, processing and interpreting information; identifying, integrating, and weighting important factors; and making trade-offs and bringing relevant factors together into a choice (Lau 2008).

The concern that an abundance of information does not guarantee informed decisions has been recognised in these fields (Hibbard and Peters 2003) and is highly relevant in the context of data-based decision-making. Everyday decisions typically concern artificially constructed information that combines only the aspects of the reality that the selected measures and signals can represent (Lau 2008). Furthermore, as only some aspects of the information are available to the decision-maker, these may include information that is not relevant in the present context and exclude information that would be essential for an informed decision (Lau 2008). Thus, the glut of information inherent in big data poses a challenge to both systems that compute and present information, and decision-makers who try to differentiate between relevant and insignificant information. Moreover, the decision-maker is seldom aware of how the parameters of the algorithms that compute the data affect the result, nor of how human cognitive processing biases the choices we make.

2.1. Data-based decision-making as a cognitive task

In this study, we are interested in a general case and a decision-making scenario in which there is large survey data or other indicators of several variables for several teams or groups of workers. In our scenario, the data are used by professionals whose task is to decide, based on the data, which one of two groups of participants is under more cognitive load and thus require support and actions. In fact, the practical motivation for our study comes from real naturalistic decision-making situations in contexts in which large data sets are utilised in organisational decision-making to improve evidence-based management; we have noticed a need to better understand the underlying cognitive phenomena in order to enhance data-based decision-making.

We will focus on the first steps in decision-making, when information is presented, and relevant information is identified. Three essential cognitive demands are in focus. First, presenting data typically involves presenting values for several attributes and subgroups. Thus, the amount of information presented and required in everyday decision-making easily exceeds human cognitive capacity. Research shows a capacity limit of about three to four items when the task requires working memory, the essential system in any cognitively demanding task that requires maintaining and processing information (Cowan 2001). This cognitive limitation constrains how humans represent and rehearse the essential features and parameters of the data, and how much of this information is used when making decisions and judgements (Allred et al. 2016). Thus, the amount of information presented is a significant factor underlying decision-making.

Furthermore, not every piece of information is relevant, and the decision-maker needs to separate the relevant information from the irrelevant, that is, to detect signals from noise, as is stressed in research on decision-making in accounting (for a review see Lau 2008). If the irrelevant information is emphasised and the problem is falsely identified, the following phases of decision-making, that is, generating alternative courses of action and selecting the most appropriate one, are based on incorrect assumptions and cannot lead to a successful result. For example, when the problem that causes poor profitability in manufacturing has not been correctly identified, the next steps in decision-making do not handle the real problem: it remains unsolved (Lau 2008).

Second, when presenting data, whether in numerical or visual format, it is seldom possible to present all the information simultaneously. Typically, several tables or figures are required to, for example, present values for different relevant attributes in two or more groups. Even if all the data are presented to us in one view, in practice, we need to focus on only one aspect at a time. Therefore, cognitive processing of the presented data requires several steps, and the task thus consists of a sequence of various cognitive sub-tasks. Processing sequential information is associated with several relevant phenomena in human cognitive functioning, including the effect of a serial order of a list item on memory recall: the first and last items in the list are typically recalled better than other items, which are the primacy and recency effects, respectively (Glanzer and Cunitz 1966).

The position of information does not only affect encoding and the memory of information, but also decision-making (Hogarth and Einhorn 1992): depending on the type of task, the items that are presented first or last during a session or task can be overemphasised and bias the judgement in a certain direction (i.e. information order bias; Perrin et al. 2001) For example, information order effects are seen as a key factor contributing to the efficiency and the effectiveness of an audit, as the information that is presented last is over-weighted in the audit (Ashton and Kennedy 2002). Similarly, in data-based decision-making, the effects related to presentation mode and organisation of information (Ashton and Ashton 1990; Yang et al. 2018) may thus affect the selection of the features that truly affect judgment.

Third, data-based decision-making is a specific example of a demanding judgment task and thus likely to be subject to the biases typical to human thinking. Research has demonstrated that our thinking seldom follows the strictly rational paths that normative decision-making theories describe; it is subject to many kinds of cognitive biases. Several biases are relevant to data-based decision-making. For example, we tend to search for evidence to match our expectations and are likely to rely on information that confirms what we want to see and find (so-called confirmation bias; Oswald and Grosjean 2004). Moreover, equivalent information can be described or 'framed' in different ways and verbalisations, which affects the choices we make (Mandel 2014; Perrin et al. 2001). Furthermore, the information presented first can frame our decisions. For example, presenting financial measures first anchors ratings so that the manager with better performance in financial measures is rated higher but if the non-financial measure is presented first, the managers who excel in these measures get higher ratings (Neumann, Roberts, and Cauvin 2011). In data-based decision-making, it is thus essential to be aware of the relevant cognitive biases that are ongoing when we look for relevant features: we may be more prone to confirming our prior expectations or our first impression of the data rather than utilising the new and unexpected information hidden within it.

2.2. Aim of the study

In our experimental study, we demonstrated cognitive phenomena that are relevant to making judgements in data-based decision-making tasks and that need attention when aiming for cognitively sound designs. We developed a type of task that represents the general cognitive demands of a data-based decision-making case and allows its systematic study. Our research questions addressed whether data-based decision-making is affected by the amount of information presented, and to what extent cognitive biases manifest in data-based decision-making.

3. Methods

The Ethics Committee of the Finnish Institute of Occupational Health has granted approval for this study.

3.1. Participants

The participants were 20 volunteers who were informed that participation required normal or corrected to normal vision, basic experience in using computers, Finnish as a native language, and no cognitive limitations or medications that could affect performance. Fourteen participants reported being women, six reported being men or other, and all but one participant reported being right-handed. The participants' age varied between 18 and 49, and the 18–29, 30–39, and 40–49 age groups had 14, 3, and 3 participants, respectively. They gave their informed consent, and received a 20 € gift card after the session.

3.2. Material

We created a new judgement task for the experiment. The task allowed manipulation of (i) the amount of presented information (length of a stimulus sequence, that is, the number of items within a sequence), (ii) the serial order of presented stimuli (position in a sequence), (iii) the difficulty of the decision-making task (unambiguous vs. ambiguous stimulus sets), and (iv) the role of background information (values for stimulus attributes defined as information irrelevant to the task).

3.2.1. Stimuli and stimulus sequences

The stimuli consisted of three critical target attributes and three noncritical background attributes associated with numbers between 1 and 10, which described the level of attribute prevalence in Group A and Group B.

A single stimulus is a three-element tuple of the form S = (attribute, score A, score B). Here, the attribute was a word, and the attribute scores were numbers between 1 and 10, describing the prevalence of the attribute for two groups, Group A and Group B. For example, the attribute 'disruptions' could have a score of 6 in Group A and a score of 8 in Group B, indicating a higher prevalence in Group B.

The attribute was one of two types: either a critical target attribute or a non-critical background attribute. Three target attributes and three background attributes were Finnish words with two to four syllables. The critical target attributes were related to workload and referred to disruptions, time pressure, and information overload (in Finnish: häiriöt, kiire, tietotulva), whereas the noncritical background attributes were neutral or related to the alleviation of workload and referred to team work, breaks and learning (tiimityö, tauot, oppiminen). The stimuli were assembled as S sequences including multiple stimuli with one, two or three critical and noncritical attributes (see Figure 1 for an example of a single stimulus and a stimulus sequence).

The sequences belonged to either of two ambiguity categories, depending on the total accumulated score for the critical target attributes in the sequence. In an unambiguous sequence, the total accumulated critical attribute in the sequence was greater for either Group A or B, that is, it was clear whether Group A or B had a higher workload. The group with the higher total accumulated score was the 'correct' choice. In an ambiguous sequence, the total accumulated critical attribute score was the same for both groups, that is, both choices were equally valid. The total accumulated score for the background attributes was always unambiguous and thus either supported or did not support either Group A or B. For example, if the background attributes supported correct response A in unambiguous sequences or the selection of Group A in ambiguous sequences, the total score for background attributes was lower for Group A than for Group B. If the background attributes

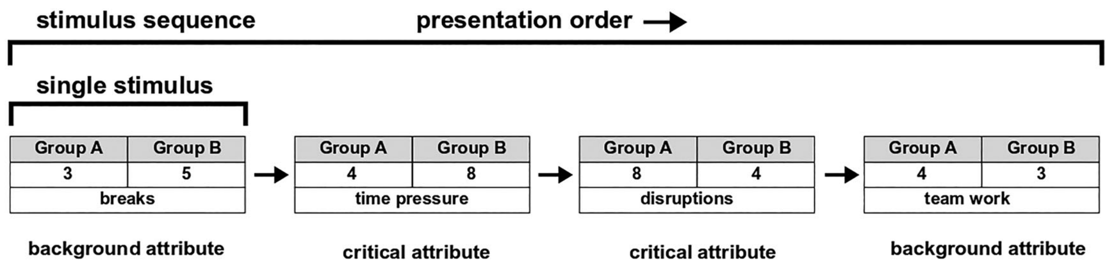

Figure 1. Example of ambiguous stimulus sequence with four single stimuli. The actual stimuli were in Finnish.

did not support correct response A in unambiguous sequences or the selection of Group A in ambiguous sequences, the total score for background attributes was higher for Group A than for Group B. The rationale was therefore that a higher total accumulated score for the critical target attributes reflected more workload, whereas a higher score for non-critical background attributes could be interpreted as reflecting less workload.

We generated sets that had (i) an equal number of stimulus sequences of length 2, 4, and 6, (ii) an equal number of unambiguous and ambiguous sequences for each sequence length, (iii) an approximately even frequency for each attribute, (iv) attributes that each appeared in all positions for each sequence length, (v) all maximum difference values (0, 1, 2, 3) present for the critical attributes, and maximum difference values (1, 2, 3) present for the background attributes. Each participant was tasked with completing at least four sets with 48 sequences each (altogether 192 sequences). A detailed description of how the ambiguous and unambiguous sequences and stimulus sets were constructed is available on request.

3.3. Procedure

The stimulus sequences were presented to the participants on a computer screen. Before each stimulus, a fixation dot was shown for 500 ms, after which the stimulus was shown for 3000 ms. After the last stimulus in the sequence, the response screen was visible until the participant pressed one of the reply keys ('F' for Group A and 'J' for Group B), or for a maximum time of 5000 ms. The participant's selection was highlighted for 500 ms. After this, the process was repeated for the next stimulus sequence (Figure 2).

We instructed the participants to determine whether Group A or Group B had a higher level of workload, based on the values for the presented critical attributes only. The response was given by pressing the 'F' key with the left-hand index finger for Group A and the 'J' key with the right-hand index finger for Group B. Participants were instructed to respond quickly and correctly. There was a brief pause after each set (48 sequences) and a 5–10-minute pause after two sets. Each of the 20 participants were tasked with completing at least four sets, and the presentation order of the sets was circularly shifted and balanced between the participants. The participants were tested individually or in groups of two to four. The session took 90–110 min and included background questions, practice trials, instructions, and a brief interview at the end of the session.

4. Results

The independent variables were (i) the amount of presented information (length of a stimulus sequence, that is, the number of items within a sequence), (ii) the

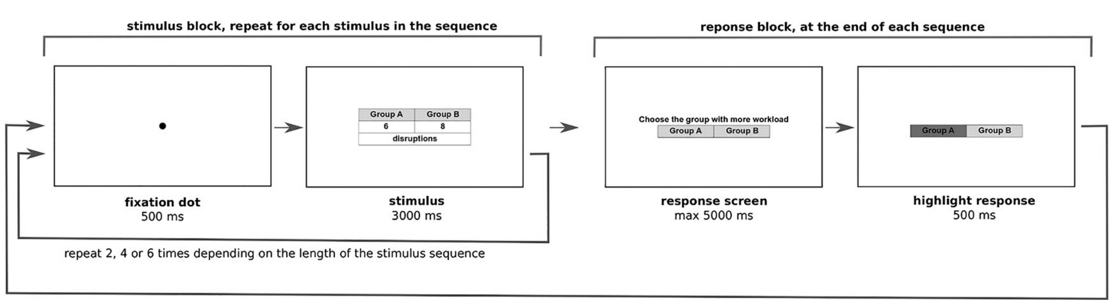

Figure 2. Presentation of stimulus sequences. The stimulus block was repeated 2, 4 or 6 times depending on the number of stimuli in the sequence. After all the stimuli in a sequence had been presented, the response block was shown. After the response block, the process was repeated for the next stimulus sequence.

difficulty of the decision-making task (unambiguous vs. ambiguous stimulus sets), and (iii) the role of background information (values for stimulus attributes defined as information irrelevant to the task). The dependent variables were the mean response time and the percentage of correct replies. We studied the effects of the Amount of Presented Information (2, 4, or 6 stimuli in a set), the Task Difficulty (unambiguous or ambiguous stimulus set), and the role of the Background Information Support (irrelevant stimulus attribute values supporting or weakening the correct response). The dependent variables were the percentage of correct answers (Figure 3 left panel), the mean response times in the judgement task (Figure 3 right panel), and the distribution of the frequency of the different responses (Table 1).

We conducted a two-way repeated-measures ANOVA to reveal the effects of Amount of Information and Background Support on the percentage of correct responses in unambiguous tasks. The Task Difficulty factor, that is, the ambiguous stimulus sets, were omitted in this analysis, as there was no correct response. The results showed the significant main effects of Amount of Information, $F(2,16) = 10.26$, $p < .01$, $\eta_p^2 = .56$ and Background Support, $F(1,17) = 5.66$, $p < .05$, $\eta_p^2 = .25$. The interaction between the main effects was not significant, $F(2,16) = 0.36$, $\eta_p^2 = .04$. The results thus show that the level of correct responses decreased as the amount of information increased. Furthermore, the level of correct responses was higher if the non-critical background information supported rather than conflicted with the correct response based on the critical target information.

We conducted a three-way repeated-measures ANOVA to reveal the effects of Amount of Information, Background Support, and Task Difficulty on the response times. All the main effects were significant, $F(2,17) = 14.84$, $p < .001$, $\eta_p^2 = .64$; $F(1,18) = 16.18$, $p < .01$, $\eta_p^2 = .47$; and $F(1,18) = 7.42$ $p < .05$, $\eta_p^2 = .29$, respectively. Furthermore, the two-way interaction effects between Amount of Information and Background Support, $F(2,17) = 4.50$, $p < .05$, $\eta_p^2 = .35$ and Task Difficulty and Background Support, $F(1,18) = 5.71$, $p < .05$, $\eta_p^2 = .24$ were significant, whereas the interaction between Amount of Information and Task Difficulty was not significant, $F(2,17) = 2.84$, $\eta_p^2 = .25$. The three-way interaction between all the main factors was significant, $F(2,17) = 4.95$, $p < .05$, $\eta_p^2 = .37$. Altogether, these results showed that response times were shorter when non-critical background information supported rather than conflicted with critical target information, and the effect of background information was stronger with more difficult tasks, that is, in ambiguous cases. The response time pattern also showed that the higher the amount of information, the shorter the response times. This result may be related to the decision-making task design: the amount of information presented and the presentation time were confounded. When the task contained more information (i.e. six attributes rather than two), the time to serially present this information also increased, offering more time to process the task before the response was required (response times were

Table 1. Frequency of actual responses A or B (Actual Response) by which response background supported stimulus sets with correct reply A or B (non-ambiguous) or either reply (ambiguous). Stimulus set length was 2.

	Background supports (actual response)				Total
Correct reply for stimulus set	A (A)	A (B)	B (A)	B (B)	
A					
Observed frequency	142_{a}	9_{b}	137_{c}	28_{b}	316
Expected values	100.0	52.1	57.4	106.6	
Residual	42.0	−43.1	79.6	−78.6	
B					
Observed frequency	21_{a}	117_{b}	13_{a}	163_{c}	314
Expected values	99.4	51.8	57.0	105.9	
Residual	−78.4	65.2	−44.0	57.1	
Either (ambiguous)					
Observed frequency	217_{a}	72_{b}	68_{b}	214_{a}	571
Expected values	180.7	94.1	103.6	192.6	
Residual	36.3	−22.1	−35.6	21.4	
Total					
Observed frequency	380	198	218	405	1201

Each subscript letter denotes a subset of categories whose column proportions did not differ significantly from each other at the .05 level.

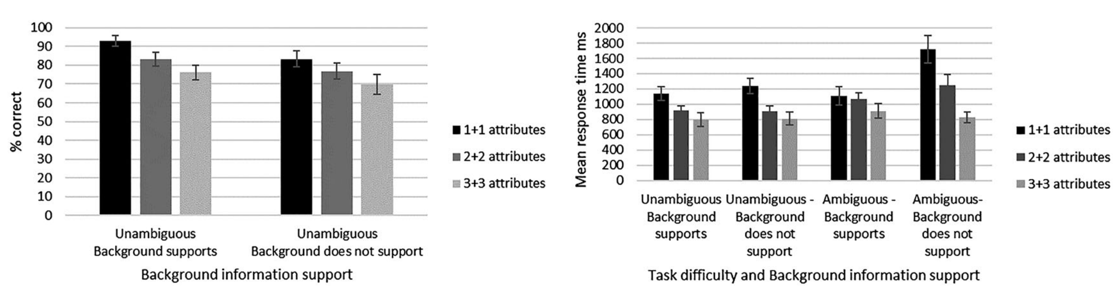

Figure 3. Percentage of correct response (left panel) and mean response time (right panel) for non-critical background information support in relation to task difficulty (unambiguous vs ambiguous critical target attributes) and amount of information (number of attributes)

calculated from the time point at which all the information was presented).

A chi-square test was conducted to reveal whether the distribution of response categories (background supports A or B * Actual response is A or B) differed within three stimulus types: correct response A (non-ambiguous sequences), correct response B (non-ambiguous sequences), or no correct response (ambiguous sequences). We analysed Post Hoc cell contributions at a level of $p < .05$ to determine whether any of the four response types was especially characteristic of some stimulus type. Table 1 presents the observed frequencies, expected values, and residuals for the response categories in stimulus set length 2. The results showed that the difference between the observed and the expected error distributions differed in all stimulus set lengths, i.e. two, four and six stimuli in a set, $\chi^2(6) = 457.7$, $p < .001$, $\chi^2(6) = 457.7$, $p < .001$, $\chi^2(6) = 457.7$, $p < .001$, respectively. Post hoc cell contributions revealed that if there was no correct reply in the task (ambiguous sequences) but the background features supported response A, participants used response category A more often than expected, and a similar pattern was evident for features supporting Group B. However, if the background attribute values did not support the specific response category, it was less frequent than expected. Moreover, this effect of background attribute values increasing the frequency of selecting the group supported by the background attribute values and decreasing the frequency of selecting the group that was not supported by background values was evident even in the distribution of responses for stimulus sets that were not ambiguous but had a correct response option.

5. Discussion

In our study, we focused on the cognitive factors relevant to making judgements in data-based decision-making tasks. The main findings were that (i) the response accuracy in data-based decision-making decreases when the amount of information increases and (ii) irrelevant information affects judgement. The results thus indicate, in the context of data-based decision-making, that cognitive demands exceeding working memory capacity impair task performance (Cowan 2001) and that judgement is subject to cognitive biases, such as being framed by inessential information (Hibbard and Peters 2003; Tversky and Kahneman 1981).

The results showed that response accuracy decreases when the amount of information increases, such as with longer stimulus sets. In our study, the fall in accuracy of judgement was clear when the number of critical and non-critical attributes increased from one to three items, that is, from two to six relevant and irrelevant attributes presented as task information. This sounds like quite a modest amount of information, and yet it overburdens human decision-makers and leads to incorrect responses. Our results indicate that a large amount of information does not automatically lead to improved judgement (Hibbard and Peters 2003), but limited human cognitive capacity is a true constraint for utilising data, as for any cognitive task requiring maintaining and processing information in the working memory (Cowan 2001).

Our results also showed that in the case of equivalent alternatives with no correct choice, participants tended to use noncritical background information in judgement. That is, they did not randomly select one of the two equally invalid response options but tended to select the option based on irrelevant attributes. This tendency was present in even non-ambiguous tasks; if an incorrect response option was selected, it tended to be the one supported by the background attribute information. Thus, in this data-based decision-making task, participants seemed to interpret irrelevant background attributes as factors that buffer against the critical attributes. This result indicates, in the context of data-based decision-making, a cognitive bias in which judgement is affected by non-essential information. It seems that irrelevant attributes can 'frame' the problem (Mandel 2014; Perrin et al. 2001) and anchor the choice (Neumann, Roberts, and Cauvin 2011). Our results thus suggest that irrelevant information in the data can lead to incorrect assumptions, hinder the handling of the real problem, and thus disrupt the selection of the appropriate further action (Lau 2008).

The contribution of our study is that we demonstrate how general cognitive limitations constrain the utilisation of data. Although the task used in this study was a simplified version of any naturalistic case, it nevertheless comprised cognitive demands relevant in any situation in which people utilise data. In our study, we thus chose to prioritise a task and method that reflect the basic cognitive processes underlying data-based decision making rather than one specific ecologically valid task. We admit that both approaches to science are essential; the dimension of the ecological validity of method, and the dimension of generalisability of the conclusions. Nevertheless, we endorse the view that when it's difficult to gain both, it is more important to ensure generalisability of the principles rather than the exact operationalisations (Banaji and Crowder 1989; Fiske and Borgida 2011). Since the methodology of our study enables us to focus on well-known and general cognitive processes operational in data-based decision-making tasks in the laboratory, we argue that the results also generalise to naturalistic contexts where these same processes are at work outside laboratories (Banaji and Crowder 1989). Therefore, the cognitive limitations demonstrated in our experiment show generalisable principles relevant to any case of data-based judgement.

Our cognitive ergonomics approach is not based on one single cognitive theory, but on various theoretical frameworks and numerous findings of cognitive psychology; this approach nevertheless advances the understanding of the factors that predict performance in cognitively complex tasks. However, in future studies, it is important to demonstrate the principles we have studied also with other data-based decision-making tasks. Furthermore, it is important to expand the external validity of our results by including cognitive and other variables outside the scope of our study. For example, the level of expertise is known to substantially affect cognitive task performance (Ericsson and Lehmann 1996). In the context of data based decision making, an interesting question would be how expertise in the subject domain, and on the other hand, in data analysis systems, would affect data-based judgement. Furthermore, there are various other relevant cognitive principles that were outside the scope of our study, but would provide useful information, such as principles of visual processing that generalise and are relevant to the design of displays presenting data.

Our aim was to understand the basic cognitive limitations and general cognitive task demands essential in the context of human-data interaction. This approach lies between earlier studies that concentrate on specific cognitive functions in a specific task and with application-driven methodology, such as the priming and anchoring effects in the context of visual analytics systems (Cho et al. 2017; Valdez, Ziefle, and Sedlmair 2018), and studies that investigate decision-making in naturalistic, knowledge-rich contexts that involve specific cognitive task demands and use a large-scale, context-related approach (Endsley 1995; Klein 2008). We claim that our approach can provide results that go beyond specific tasks and specific contexts. In the future, it is important to develop theories for applied cognitive psychology that address how the human cognitive system is able to perform well in any cognitively demanding task (Logie, Trawley, and Law 2011); for example, how the unified human cognitive system functions, with all its limitations and various information processing components, when we efficiently utilise data in decision-making tasks.

5.1. Cognitive ergonomics for data analyses

The cognitive phenomena demonstrated in our experimental study suggest principles that require attention when we aim to improve the quality of data-based decisions. Advances in digitalisation and computation can also provide new means that enable designing cognition friendly systems. Human-data interaction happens when people use applications to find meaningful information in the data. It would be useful to utilise data on human behaviour in this context in a similar way to which the data obtained when people use search engines to navigate in the internet is utilised. In human-data interaction, technical systems can help identify specific limitations and compensate for the limited human cognitive capacities that impair our ability to efficiently utilise data. They can also adapt to users' biases and present and target data in a way that supports informed choice (Valdez, Ziefle, and Sedlmair 2018).

First, when designing systems, we need to find ways in which to organise information and reduce users' cognitive burden (Hibbard and Peters 2003). A good cognitive ergonomic principle is to support the encoding of a valid picture of the situation by, for example, shortening the sequences when presenting information or by supporting representation construction if several stages and sub-processes are required when interpreting the data (Hibbard and Peters 2003; Yang et al. 2018). A clever sequential presentation system would decrease cognitive load in situations that require the creation of a mental model of a vast amount of data. Moreover, since the presentation order of information matters, technical systems can model its effect and weigh the information that is underrepresented. Thus, a technical system would be able to computationally correct the human tendency to rely more on information that is presented first or last.

Second, cognitively sound systems should communicate to the user those aspects of the data that are meaningful in the specific context. On the one hand, the system should be able to help the user identify and weigh the information that is relevant to the choice by, for example, drawing attention to specific attributes or outcomes by framing them (Hibbard and Peters 2003). On the other hand, the system could make human decision-makers aware that their decisions may be framed by incidental attributes which lead them to lean towards a decision that is based on unimportant aspects of the data.

Human–data interaction systems provide objective behavioural data that can be more useful in enhancing decision-making than subjective user preferences which do not necessarily correlate with how well the information can be utilised in decision-making (Hibbard and Peters 2003). Results concerning response times and correct responses in well-defined contexts and controlled tasks, such as our experiment, can provide reference patterns for comparison when interpreting data obtained in more complex cases. For example, our results showed longer response times in ambiguous cases when the selected response conflicted with the non-critical background information. This result suggests that selecting a response even if the situation is unclear and a general response pattern would predict another response, may be comparable to incorrect choices. In behavioural sciences, a typical response and reaction time pattern shows longer responses for incorrect than for correct decisions. Using behavioural data to identify 'outliers' in response patterns can be developed for a technical solution that recognises when human decision-makers need more support to improve the quality of their decisions.

In sum, good cognitive ergonomics of design should at least (i) support the encoding of the reality of the situation by, for example, shortening the sequences when presenting information; (ii) enhance representation construction if several stages and subprocesses are required when presenting the data using, for example, a clever sequential presentation system; (iii) support ways of communicating which aspects of the data are considered meaningful to the user in the specific context; and (iv) make the human decision-maker aware of incidental attributes that may frame their decisions and cause them to lean towards unimportant aspects of the data.

We hope that our cognitive ergonomics approach will inspire future research. The results of our study need to be expanded in order to demonstrate the variety of cognitive demands relevant to data-based decision-making contexts. Research and design should utilise the vast amount of knowledge cognitive psychology already provides and explicitly define how the general limitations of the human cognitive system manifest when humans utilise data. With evidence-based and cognitively sound design, human–data interaction systems can develop information presentation techniques that facilitate encoding and utilising a larger amount of data than that which would be possible for the limited human cognitive system without support.

Acknowledgments

This study was supported by the Academy of Finland (decisions 326280 and 326339). We are thankful to Mr. Tomi Passi for his help in recruiting participants and running the experiments.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by Academy of Finland [grant number 326280, 326339].

References

Allred, S. R., L. E. Crawford, S. Duffy, and J. Smith. 2016. "Working Memory and Spatial Judgments: Cognitive Load Increases the Central Tendency Bias." *Psychonomic Bulletin & Review* 23 (6): 1825–1831. doi:10.3758/s13423-016-1039-0

Ashton, R. H., and A. H. Ashton. 1990. "Evidence-responsiveness in Professional Judgment – Effects of Positive Versus Negative Evidence and Presentation Mode." *Organizational Behavior and Human Decision Processes* 46 (1): 1–19. doi:10.1016/0749-5978(90)90019-6.

Ashton, R. H., and J. Kennedy. 2002. "Eliminating Recency with Self-Review: The Case of Auditors' 'Going Concern' Judgments." *Journal of Behavioral Decision Making* 15 (3): 221–231. doi:10.1002/bdm.412.

Baddeley, A. D., and G. Hitch. 1974. "Working Memory." In *Vol. 8. The Psychology of Learning and Motivation*, 47–89. New York: Academic Press.

Banaji, M. R., and R. G. Crowder. 1989. "The Bankruptcy of Everyday Memory." *American Psychologist* 44 (9): 1185–1193. doi:10.1037/0003-066x.44.9.1185.

Cho, I., R. Wesslen, A. Karduni, S. Santhanam, S. Shaikh, and W. Dou. 2017. "The Anchoring Effect in Decision-Making with Visual Analytics." IEEE Conference on Visual Analytics Science and Technology (VAST), Phoenix, AZ, 2017, pp. 116–126. doi:10.1109/VAST.2017.8585665.

Cowan, N. 2001. "The Magical Number 4 in Short-Term Memory: A Reconsideration of Mental Storage Capacity." *Behavioral and Brain Sciences* 24 (3): 87–114.

Endsley, M. R. 1995. "Toward a Theory of Situation Awareness in Dynamic-Systems." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37 (1): 32–64. doi:10.1518/001872095779049543.

Endsley, M. R., and M. D. Rodgers. 1994. "Situation Awareness Information Requirements Analysis for En Route Air Traffic Control." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 38 (1): 71–75.

Ericsson, K. A., and A. C. Lehmann. 1996. "Expert and Exceptional Performance: Evidence of Maximal Adaptation to Task Constraints." *Annual Review of Psychology* 47: 273–305. doi:10.1146/annurev.psych.47.1.273.

Fiske, S. T., and E. Borgida. 2011. "Best Practices: How to Evaluate Psychological Science for Use by Organizations." *Research in Organizational Behavior* 31: 253–275. doi: 10.1016/j.riob.2011.10.003.

Glanzer, M., and A. R. Cunitz. 1966. "Two Storage Mechanisms in Free Recall." *Journal of Verbal Learning and Verbal Behavior* 5: 351–360.

HFES (Human Factors and Ergonomics Society). Accessed February 1, 2019. https://www.hfes.org/home.

Hibbard, J. H., and E. Peters. 2003. "Supporting Informed Consumer Health Care Decisions: Data Presentation Approaches That Facilitate the Use of Information in Choice." *Annual Review of Public Health* 24: 413–433. doi:10.1146/annurev.publhealth.24.100901.141005.

Hogarth, R. M., and H. J. Einhorn. 1992. "Order Effects in Belief Updating - the Belief-Adjustment Model." *Cognitive Psychology* 24 (1): 1–55. doi:10.1016/0010-0285(92)90002-J.

IEA (International Ergonomics Association). Accessed February 1, 2019. https://www.iea.cc/.

Kalakoski, V. 2016. *Cognitive Ergonomics*. Accessed February 1, 2019. https://oshwiki.eu/index.php?title=Cognitive_ergonomics&oldid=245803.

Klein, G. 2008. "Naturalistic Decision Making." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 50 (3): 456–460. doi:10.1518/001872008x288385.

Klein, G. A., J. Orasanu, R. Calderwood, and C. E. Zsambok. 1993. "Preface." In *Decision Making in Action: Models and Methods*, edited by G. A. Klein, J. Orasanu, R. Calderwood, and C. E. Zsambok, vii–vix. Westport, CT: Ablex Publishing.

Lau, Y. W. 2008. "Noise in Accounting Information: The Signal Detection Theory Perspective." *International Journal of Economics and Management* 2 (1): 179–193.

Logie, R. H., S. Trawley, and A. Law. 2011. "Multitasking: Multiple, Domain-Specific Cognitive Functions in a Virtual Environment." *Memory & Cognition* 39 (8): 1561–1574. doi:10.3758/s13421-011-0120-1.

Mandel, D. R. 2014. "Do Framing Effects Reveal Irrational Choice?" *Journal of Experimental Psychology-General* 143 (3): 1185–1198. doi:10.1037/a0034207.

Neumann, B. R., M. L. Roberts, and E. Cauvin. 2011. "Stakeholder Value Disclosures: Anchoring on Primacy and Importance of Financial and Nonfinancial Performance Measures." *Review of Managerial Science* 5 (2-3): 195–212. doi:10.1007/s11846-010-0054-1.

Oswald, M. E., and S. Grosjean. 2004. "Confirmation Bias." In *Cognitive Illusions. A Handbook on Fallacies and Biases in Thinking, Judgement and Memory*, edited by R. F. Pohl, 79–96. Hove and NY: Psychology Press.

Pariser, E. 2011. *The Filter Bubble: What the Internet is Hiding From you*. New York, NY: Penguin.

Perrin, B. M., B. J. Barnett, L. Walrath, and J. D. Grossman. 2001. "Information Order and Outcome Framing: An Assessment of Judgment Bias in a Naturalistic Decision-Making Context." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 43 (2): 227–238. doi:10.1518/001872001775900968.

Tversky, A., and D. Kahneman. 1981. "The Framing of Decisions and the Psychology of Choice." *Science* 211 (4481): 453–458. doi:10.1126/science.7455683.

Valdez, A. C., M. Ziefle, and M. Sedlmair. 2018. "Priming and Anchoring Effects in Visualization." *IEEE Transactions on Visualization and Computer Graphics* 24 (1): 584–594.

Wickens, C. D., S. E. Gordon, and Y. Liu. 2004. *An Introduction to Human Factors Engineering*. 2nd ed. Upper Saddle River, NJ: Pearson Prentice Hall.

Yang, W. T., J. H. Ma, H. Z. Chen, A. G. Maglione, E. Modica, D. Rossi, and F. Babiloni. 2018. "Good News or Bad News, Which do you Want First? The Importance of the Sequence and Organization of Information for Financial Decision-Making: A Neuro-Electrical Imaging Study." *Frontiers in Human Neuroscience* 12. doi:10.3389/fnhum.2018.00294.